The First Step Towards a System of Open Digital Scholarly Communication Infrastructure

David W. Lewis, Mike Roy, and Katherine Skinner

A guest post for *IO: In the Open, December 11, 2018*

We are working on a project to map the infrastructure required to support digital scholarly communications. This project is an outgrowth of David W. Lewis' "2.5% Commitment" proposal.

Even in the early stages of this effort we have had to confront several uncomfortable truths.

First Uncomfortable Truth: In the main, there are two sets of actors developing systems and services to support digital scholarly communication. The first of these are the large commercial publishers, most notably Elsevier, Wiley, and Springer/Nature. Alejandro Posada and George Chen have documented their efforts. A forthcoming SPARC report previewed in a DuraSpace webinar by Heather Joseph confirms these findings. The second set of actors may currently be more accurately described as a ragtag band of actors: open source projects of various sizes and capacities. Some are housed in universities like the Public Knowledge Project (PKP), some are free standing 503(C)3s, and others are part of an umbrella organization like DuraSpace or the Collaborative Knowledge Foundation (COKO). Some have large installed bases and world-wide developer communities like DSpace. Others have yet to establish themselves, and do not yet have a fully functional robust product. Some are only funded with a start-up grants with no model for sustainability and others have a solid funding based on memberships or the sale of services. This feels to us a bit like the rebel alliance versus the empire and the death star.

Second Uncomfortable Truth: The commercial players are well resourced and are building or acquiring tools that can be integrated into an end-to-end scholarly and research workflow. The ragtag band of open source projects are underfunded and the tools they produce are at best only loosely integrated. Despite some early effort, particularly by Bianca Kramer and Jeroen Bosman, there is currently no comprehensive view of the landscape (thus our project). We therefore do not know what is required, or where there are gaps and redundancies, for an open community controlled end-to-end workflow. While most projects adhere to accepted standards, efforts at integration are in their infancy. The work of the Joint Roadmap for Open Science Tools (JROST) is a good beginning, but it is only just starting. Importantly because we have only a vague picture of the open infrastructure community, we don't know what it will cost to build what is required. This means it is nearly impossible to build a case for system-wide funding, especially one based on ROI.

Third Uncomfortable Truth: There is little transparency anywhere. The large commercial providers have little interest in disclosing the particulars of their strategies, as they consider them trade secrets. Unfortunately, many open projects are not transparent either. In part because open projects have such a mix of organizational structures it is hard to find financial or even governance information about projects. It is thus hard for those who wish to contribute to projects to understand whether or not they are making good investments.

These uncomfortable truths lead us to make several suggestions to the open community. We believe that absent a significant change in the way in which the open community functions it will have little chance to successfully compete against the large commercial providers.

We believe the first required change is for open projects to become more transparent. In the current environment, as the recent failure of the Digital Preservation Network reminds us, too many projects fail, either a slow death because their financial support is inadequate to maintain their systems over time, or because their grant ends before a market ready product is delivered, or because they simply have no path to financial sustainability and just run out of money. Sometimes the project was just too technically ambitious or they have no capacity or understanding of how to market what they have built. We need a way to easily and reliably gauge the technical, organizational, and financial health of projects in order to assure investors that they are making wise use of their money.

We believe a lightweight, standard audit process for all open projects could help. Geoffrey Bilder, Jennifer Lin, and Cameron Neylon have proposed principles that can be used to create such a reporting mechanism. The Global Sustainability Coalition for Open Science Services (SCOSS) has developed a process for analyzing open projects, though to date it has only been used in a limited number of cases. Beyond financial information, a standard means of assessing the organizational capacity would demonstrate the health of the project's governance structure. Educopia's *Community Cultivation: A Field Guide* can serve as the basis for this. We need mechanisms for judging the fiscal, technical, and organizational health of projects. We think of this as providing the basis for an "Angie's List" for digital scholarly communications infrastructure. This will provide the community with the confidence necessary to make the needed investments to build infrastructure that is competitive with the commercial sector. If we do not do so, the only choice will be to take what is offered by the large for-profit players.

The second required change is to develop a map of the landscape. Our project intends to do this. The map will document the existing infrastructure and show gaps and redundancies. This is the first step in developing strategies for integrating systems across the whole of the scholarly and research workflow. Together with financial and organizational information, it will then be possible to create a roadmap for the development and maintenance of the full range of infrastructure that is required. From this it will be possible to develop a case to fund an alternative, community-owned and community-led system of infrastructure. We believe our roadmap will provide government agencies and foundations, as well as to universities and their libraries the confidence necessary to make significant and ongoing contributions to open systems.

Our project will provide some preliminary data, but regular and consistent collection of data from projects will be required, as will the maintenance of a continually revised roadmap. This will require an organizational and governance structure that is trusted by the community. Such an organization does not exist today. We need to start to thinking about how to create it.

As we begin to look at the projects that currently make up the open infrastructure for digital scholarly communication we believe there is much that the community can be proud of. But it is also quite clear that what we now have is far from what is required. We believe that a critical first step is to create a more transparent environment where it is easy to judge the financial, technical, and organizational capacity of projects. Without this transparency it will be impossible to build the trust necessary for

projects to collaborate and integrate, and the trust that is required for funders to make the necessary investments.

We believe that developing the means of transparency is the required first step and that now is time to take it.